

**METHOD FOR PROVIDING A COMPRESSED RENDITION OF A
VIDEO PROGRAM IN A FORMAT SUITABLE FOR ELECTRONIC
SEARCHING AND RETRIEVAL**

Spa

TECHNICAL FIELD

5 This invention relates generally to a method for automatically providing a compressed rendition of a video program in a format suitable for electronic searching and retrieval, and more particularly to a method for providing a compressed rendition of a video program in a format suitable for electronic searching and retrieval on the World Wide Web.

10 **BACKGROUND**

004270" 6035960
09525039 072400

The rapid growth of the World Wide Web began with the development of an on-line browser having a graphical user interface. Graphical interfaces provide a number of important advantages, including the ability to rapidly scroll through a document to get to a particular point of interest. Moreover, the ability to interact with a medium other than text (i.e. images or audio) increases the rate at which information can be conveyed since an image often conveys an idea faster and more efficiently than text.

15 While graphical browsers provide an adequate interface for text and images, they provide an inadequate interface for video programs. The sequential nature of the video and audio components of a video program impedes rapid access to such programs on the World Wide Web by graphical browsers.

20 Furthermore, because of the limited bandwidth of networks supporting the World Wide Web, and particularly the limitations of most users' connections to such networks, it takes a long time to transmit a program with its full content.

25 For example, at a connection speed of 28,800 bits per second, it could take up to about 45 minutes to transmit even a three or four minute audiovisual segment

with sound and full-motion video. As a result, video program providers sometimes form a compressed version of the video program by manually extracting and retaining selected frames from the program while other frames are discarded. The selected frames and accompanying text, typically taken from a transcript of the program, result in a document that may subsequently be made available over the World Wide Web. However, the generation of this document is typically a tedious and time consuming task since it must be created by a manual process.

Accordingly, it would be advantageous to provide a rendition of a video program which can be automatically generated and which allows easy interaction with graphical browsers with a minimum of information loss.

SUMMARY OF THE INVENTION

The present inventors have realized that a pictorial transcript representation of a video program is particularly well suited for on-line searching and retrieving applications such as browsing on the World Wide Web. Pictorial transcripts are compact representations of video programs which are automatically generated by selecting representative frames or images from the video program and combining them with a second media component such as audio or text which is associated with each representative frame. Properly chosen, the representative frames convey a substantial portion of the information content of the original video program. Moreover, pictorial transcripts may be generated in an automatic fashion, thus eliminating the substantial time and effort that was previously required to place a document of this type on the World Wide Web.

The inventive method provides a compressed rendition of a video program in a format suitable for electronic searching and retrieval. An electronic pictorial transcript representation of the video program is initially received. The video program has a video component and a second information-bearing media

component associated therewith. The pictorial transcript representation includes a representative frame from each segment of the video component of the video program and a portion of the second media component associated with the segment. The electronic pictorial transcript is transformed into a hypertext format to form a hypertext pictorial transcript. The hypertext pictorial transcript is subsequently recorded in an electronic medium.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is an example of one page of a printed pictorial transcript generated from a television news program in accordance with method of the present invention.

FIG. 2 illustrates the use of server push for viewing an HTML pictorial transcript.

FIG. 3 shows an example of a page format that may be employed when performing keyword searching.

FIG. 4 shows an example of an index that may be generated for HTML pictorial transcripts.

DETAILED DESCRIPTION

A method for automatically compressing multimedia data is disclosed in U.S. Patent Application Serial No. 08/252,861 (U.S. Patent No. 6,271,892) filed June 2, 1994, and Shahraray B., and Gibbon D.C., "Automatic Generation of Pictorial Transcripts of Video Programs," in *Multimedia Computing and Networking* 1995, Proc. SPIE 2417, February 1995, the latter reference being hereby incorporated by reference. In accordance with this known method, a video program is compressed by selecting certain frames from the entire sequence of frames to serve as representative frames. For example, a single frame may be used to represent the visual information contained in any given scene of the video program. A scene may be defined as a segment of the video program over which the visual contents do not

change significantly. Thus, a frame selected from the scene may be used to represent the entire scene without losing a substantially large amount of information. A series of such representative frames from all the scenes in the video program provides a reasonably accurate representation of the entire video program with an acceptable degree of information loss. These compression methods in effect perform a content-based sampling of the video program. Additional information may be found in B. Shahraray, "Scene Change Detection and Content-Based Sampling of Video Sequences," *Digital Video Compression: Algorithms and Technologies* 1995, SPIE 2419.

10 In the previously cited documents, a plurality of representative frames are selected by sampling the video program in a content-based manner to retain a single representative frame from each scene. While the series of frames selected in this manner may not contain all the visual information in the original video program, when combined with another medium that was a part of the original video program, such as audio or closed-captioned text, the resulting multimedia program adequately conveys the information content of the video program in a condensed format. To generate this condensed multimedia program, a correspondence must be formed between the representative frames and the audio or textual medium. For example, each representative frame should be associated with the portion of the audio or textual medium corresponding to the entire scene from which the representative frame was selected. This correspondence may be accomplished in a relatively simple manner because in the original video program the video medium is already synchronized with the audio or textual information. Additional details concerning the formulation of this correspondence may be found in the previously cited references.

The representative frames, the audio or textual components associated therewith, and the correspondence between the representative frames and the

audio or textual components comprise electronic data representing a condensed version of a video program, which hereinafter will be referred to as the condensed electronic data.

In the case of closed-captioned text, a printed rendition of the condensed electronic data may be provided. The printed rendition constitutes a so-called pictorial transcript in which each representative frame is printed with a caption containing the portion of the closed-caption text corresponding to the scene from which that representative frame is taken. FIG. 1 is an example of one page of printed pictorial transcript generated from a television news program.

Alternatively, rather than printing the condensed electronic data as a pictorial transcript, the data simply may be electronically stored for subsequent retrieval. Thereafter the data may be printed, displayed on a computer, or transmitted in any desired format.

In addition, the condensed electronic data may be generalized further to refer to the series of representative frames and the audio segments corresponding thereto rather than closed-caption segments. In this case the condensed electronic data may be conveniently stored electronically and then displayed by sequentially displaying the representative frames and, simultaneous with each displayed frame, playing the corresponding audio segment.

In accordance with the present invention, electronic data representing a condensed version of a video program is formatted in hypertext markup language (HTML) so that the resulting HTML document is compatible with the World Wide Web. HTML documents refer to on-line documents having words or graphics that contain links to other on-line documents. Such documents are commonly referred to as hypertext documents. By selecting the link (using a mouse or key command) the user is connected to another document that may be

located on the same or a different computer. It should be noted that while the present invention is described in terms of an on-line document formatted in HTML, more generally the present invention is applicable to hypertext documents formatted in languages other than HTML, such as hypercard, for example.

An HTML document is automatically produced from the condensed electronic data by an HTML generator, which converts the data into an HTML document. Procedures to implement such a generator are well known. As used hereinafter, the terms HTML document and HTML pictorial transcript refer to the condensed electronic data that is formatted in HTML. The HTML document or pictorial transcript may be composed of individual records connected by links. The individual records of the HTML document or pictorial transcript are referred to as pages.

The HTML pictorial transcript may be advantageously divided over two or more HTML pages, depending on the size of the document. An HTML document consisting of only a single HTML page is impractical for all but the shortest programs (e.g., less than ten minutes in length) because WWW browsers, which sometimes lack parallel loading capability, begin to exhibit unacceptable delays. In fact, even browsers having parallel loading capability such as Netscape will often be taxed. The size of each HTML page may be determined in any appropriate manner. For example, the HTML generator may begin a new page after a predetermined number of images (e.g., 25) have been placed on a single page. Alternatively, the pages may be divided on the basis of story and topic based segmentation. The various pages comprising the HTML document may connected by hypertext links.

A graphical browser is a graphical interface that can access documents on the WWW in an HTML format. The HTML pictorial transcript may be

conveniently accessed and searched using conventional graphical browsers such as Mosaic, Spry and Explorer, for example.

The HTML pictorial transcript may be displayed in a variety of different formats. The user may have the option of selecting among several
5 predetermined formats, or alternatively, the user may customize a format via the web browser. The server, in turn, re-executes the HTML generator routine, which now produces the HTML document in the desired format. Additionally, if no selection is made, the HTML transcript may be displayed in a default format (which may be one of the standard formats). In some embodiments of the
10 invention, the user may be provided with a plurality of different default formats from which to choose.

In one embodiment of the invention, a standard or default format displays an HTML pictorial transcript that is the equivalent of the printed rendition of a pictorial transcript such as shown in FIG. 1. Other formats may modify this
15 particular format to reduce retrieval time and improve page layout. For example, some formats may be employed to reduce the required bandwidth by displaying only a subset of the representative frames contained in the HTML pictorial transcript. Many different criteria may be employed to determine which representative frames to retain and which to omit.

One criterion that may be used to eliminate select representative frames is
20 based on the presence of redundant frames. For example, if the original program contains a shot of a given scene at one time and subsequently contains substantially the same scene after one or more other scenes have intervened, the resulting pictorial transcript will contain two representative frames that are
25 substantially the same. Accordingly, one of the redundant representative frames may be eliminated to reduce bandwidth. In the resulting HTML pictorial transcript it may be desirable to use a hypertext link in place of the second

Other criteria that may be used to eliminate select representative frames are based on random subsampling (e.g., retain every other representative frame) or, alternatively, the size of the JPEG image file. For example, it may be desirable to retain only the largest of the image files on the assumption that image size is correlated with the complexity of the image. More complex images typically convey more information. Conversely, it may be desirable to retain only the smallest of the image files to further minimize bandwidth requirements.

15 Another criterion that may be employed to select a subset of the
representative images is based on the length of the scene from which the
representative image was taken. For example, only representative images taken
from the longest of the scenes in the video program may be retained since these
scenes are presumably the most significant. For example, a video program of a
20 speaker making a presentation before an audience may contain many longer
scenes of the speaker interrupted by occasional brief shots of the audience. If the
representative frames from only the longest scenes are retained, then
representative frames of the speaker will be retained while the representative
frames of the audience will be eliminated.

8

removed because most commercials are either not captioned or are captioned in a mode different from the remainder of the video program. Accordingly, the change in caption modes can be used to detect advertisements which are to be omitted from the HTML transcript.

5 Another format that may be used to display HTML pictorial transcripts takes advantage of a mechanism known as server push, which is available on recent versions of the Netscape browser. Server push allows an HTML page to undergo changes while it is being viewed. This browser feature can be used to maintain a suitable page layout (e.g., a layout having a maximum number of
10 images) without needing to eliminate sequentially occurring images. This feature, which could also be implemented using Java Animations, will be illustrated with reference to FIG. 2. FIG. 2(a) shows an HTML page of a pictorial transcript which has three sequential images 1, 2, and 3, without any intervening captions. However, suppose the page format which is selected dictates that only
15 one image is to be displayed on a page, as in FIG. 2(b). Server push may be used display the images as shown in FIGS. 2(c) - 2(e). When the page is first displayed at time t_1 in FIG. 2(c), only the first image is displayed. Using server push, the second image can be displayed at a later time t_2 (e.g. one second later), as shown in FIG. 2(d). At yet a later time t_3 the third image can be displayed, as in FIG.
20 2(e). Moreover, if the network bandwidth and client and server throughput are sufficiently high, video shorts (real-time playback) can be made to appear at the caption breaks.

In many cases a user will not be interested in viewing the HTML pictorial transcript in a sequential manner. Rather, the user may be only interested in
25 those portions of the transcript that pertain to a particular topic. In such cases the user may wish to perform a keyword search of the HTML pictorial transcript. The HTML generator can perform the search on the closed-captioned text and

emphasize those portions of the transcript that contain the keyword. For example, images that appear immediately prior to and after the occurrence of a keyword may be displayed at full resolution while other images may be displayed at a smaller size and resolution. FIG. 3 shows an example of this format after a search for the word "Tokyo." The smaller images may be hypertext links to the corresponding full sized images. In some cases, particularly for large HTML pictorial transcripts, hypertext anchors may be used in place of the small images to reduce bandwidth. If the keyword appears more than once in the transcript, a chain of links may be created among the individual occurrences of the word. For example, in FIG. 3, the arrows denote a link to other occurrences of the term "Tokyo." The HTML pictorial transcript may also include hypertext anchors to other HTML documents which contain material supplementary to, or related to, the information in the transcript.

The HTML generator may create an index page for the HTML pictorial transcript using conventional methods such as linguistic techniques, for example.. FIG. 4 shows one example of such an index page, which may be located as the first page of the document. The index may contain links to the individual pages of the transcript. The index may also include other information such as index terms obtained by linguistic analysis techniques. In FIG. 4, a portion of the index is available for the user to list additional keywords to serve as index terms. The index terms may be hypertext links to those locations in the transcript where the terms appear.

Similar to the HTML documents previously discussed, HTML pictorial transcripts in which the representative frames are each associated with a corresponding audio segment may be arranged in a variety of different formats. For example, the individual representative frames may serve as links to the

audio segment. Alternatively, anchors may be associated with the representative frames. By clicking on the anchors the respective audio segments are played.

It will be appreciated that those skilled in the art will be able to devise numerous arrangements which, although not explicitly shown or described
5 herein, embody the principles of the invention. Accordingly, all such alternatives, modifications and variations which fall within the spirit and broad scope of the appended claims will be embraced by the principles of the invention. For example, while the invention has been described as electronic data representing a condensed version of a video program that is formatted as an
10 HTML document for the World Wide Web, the invention is more generally applicable to such data that is formatted in any hypertext language suitable for electronic retrieval on a computer or over a communications network.

004220-6805260